ABSTRACT

Viruses are obligate parasites that have proteinaceous capsids enclosing the genetic material X-ray diffraction studies on single crystals of viruses enable visualization of the structures of intact virus particles at near-atomic resolution These studies provide detailed information regarding the coat protein folding, capsid architecture, molecular interactions between protein subunits, plausible sites of receptor recognition, role of metal ions in the capsid structure and assembly and disassembly The present thesis deals with the structure solution and analysis of coat protein (CP) mutants from *Sesbania mosaic virus* (SeMV)

SeMV infects *Sesbania grandiflora* belonging to *Fabaceae* and is isolated from farmer's fields in Tirupathi, Andhra Pradesh, India. It is a single-stranded positive sense RNA virus with a genome length of 4149 nucleotides The genome encodes four potential overlapping open reading frames (ORFs) ORF1 codes for an 18 kDa protein that is proposed to be involved in the movement of the virus The ORF2 encodes a 105 kDa protein believed to be a polyprotein (single polypeptide chain encompassing more than one functional protein) The ORF3, internal to ORF2 is proposed to be expressed as a *trans*-frame polyprotein, translation brought about by ribosomal frameshifting mechanism Coat protein (CP) is encoded by the fourth ORF present at the 3' end of the genome (Lokesh et al , 2001) CP is responsible for the encapsidation and protection of the viral genome

The structure of SeMV determined at 3Å resolution showed that the CP adopts a jellyroll β-sandwich fold with eight antiparallel β-strands connected by few short helices and loops The asymmetric unit of the capsid is composed of chemically identical A, B and C subunits arranged in quasi-equivalent environments and four ion binding sites The A type subunits form pentamers while the B and C type subunits form hexamers in the T=3 capsids The amino terminal arms of the CP are ordered from residue 46 in the C subunits while in A and B subunits they are ordered from residue 73 The ordered amino terminal arms of C type subunits form a β-annulus like structure at the quasi 6-fold axes (residues 48-58) The β-annulus and dimeric interactions form a continuous scaffold connecting all the C subunits and the calcium

ion imparts stability to the capsid by charge compensation Calcium ions are present at the intersubunit interfaces around the quasi 3-fold axes The release of calcium ions inside the plant cell is believed to be an early step in the disassembly of the native virus Apart from these, an important motif is the "N-ARM" at the amino terminus from residues 28-36 that consists of a basic stretch of 7 arginines The presence of an N-terminal basic segment is a feature shared by all the members of the sobmovirdae family It is of interest to investigate the role of these structural features in the assembly and stability of the native virus

Earlier, the expression of CP gene in *E coli* was shown to result in the assembly of T=3 virus-like particles (VLPs) that were found to encapsidate CP mRNA and *E coli* 23S rRNA Further, it was shown that the deletion of N-terminal 65 amino acid residues resulted in the formation of T=1 particles (CP-NΔ65) Deletion of 36 residues led to the assembly of pesudo T=2 and T=1 particles (CP-NΔ36) N-terminal 22 amino acids were dispensable for T=3 capsid formation (CP-NΔ22) Recent mutational studies have shown that deletion of 31 residues from the amino terminus also forms T=1 capsids (CP-NΔ31) Further, the substitution of a few (CP-R32-36E) or all (CP-R28-R36E) Arg to Glu in the ARM led to less stable, empty T=3 capsids (CP-R28-36E) In another assembly study, the residues forming the β-annulus were partially (CP-Δ48-53) or fully deleted (CP-Δ48-58) and it was established that these residues are dispensable for T=3 capsid assembly Mutational studies with calcium ligands in rCP (CP-CΔ2) and CP-NΔ65 (CP-NΔ65-D146N-D149N and CP-NΔ65-CΔ2) suggested that the stability of the capsids is drastically reduced in the absence of calcium ions

A review of current literature on the available crystal structures of viruses and the interactions involved in capsid assembly are presented in chapter I of the thesis The experimental procedures employed in the study such as protein expression and purification, crystallization, X-ray data collection, data processing, structure solution, model building, refinement and structure analysis of different mutants used in this study etc , are described in detail in chapter II

Chapter III deals with the structural studies on recombinant and mutant T=3 capsids of SeMV The details of steps undertaken from crystallization to structural analysis for rCP, CP-NΔ22, CP-P53A and crystallization of CP-CΔ2, CP-Δ48-58, CP-Δ48-53, CP-R28-36E and CP-R32-36E are presented in this chapter The rCP and CP-NΔ22 capsids crystallized in the spacegroup R3 and CP-P53A capsids crystallized in the spacegroup C2 X-ray diffraction data to 3 6Å, 4 1Å and 5 5Å resolution were collected from single crystals of rCP, CP-P53A and CP-NΔ22, respectively, at liquid nitrogen temperature Crystals of CP-CΔ2, CP-Δ48-58, CP-Δ48-53, CP-R28-36E and CP-R32-36E did not diffract X-rays The orientation of the particle in the unit cell was determined by locked self-rotation functions The structure solution of rCP was achieved using molecular replacement starting with the native SeMV capsid as the phasing model Electron density NCS averaging was carried out starting with phases obtained from a polyalanine model of the native capsid and the structure factor amplitudes of rCP At the end of averaging, the R factor and the correlation coefficient were 28 7% and 73 8%, respectively (R defined as $100^{*}\ \Sigma(|F_o-F_c|)/\Sigma F_o$ and correlation coefficient as $\Sigma\ ((F_o-\langle F_o\rangle)^{*}(F_c-\langle F_c\rangle)/(\Sigma(F_o-\langle F_o\rangle)^2)^{*}(\Sigma(F_c-\langle F_c\rangle)^2)^{1/2})$ Initial structure solutions of both CP-P53A and CP-NΔ22 were obtained using the molecular replacement program AMoRe The correlation coefficient and R values at the end of AMoRe structure solution were 49 9% and 34 9% for CP-P53A and 48 1% and 38 5% for CP-NΔ22 The refinement of rCP, CP-P53A and CP-NΔ22 were done using CNS *version* 1 1 and the polypeptide models for rCP and CP-P53A were built into the final electron-density map using the interactive graphics program O

The quality of the map was good enough for building the model and unambiguous positioning of the side chains The C subunit is traced from residue 44 in rCP and residue 46 in CP-P53A A and B subunits are traced from residues 73 and 72, respectively, in both rCP and CP-P53A The rCP and CP-P53A superpose well with the native CP with an overall rmsd of 0 3Å The intersubunit interactions in recombinant capsids match well with those of the native capsids The presence of a calcium ion at the intersubunit interfaces of the A, B and C subunits is a feature common to native as well as recombinant SeMV capsids As in the native capsid, the

calcium adopts an octahedral coordination and only five out of six ligands are traced in the electron density map At the quasi 6-fold axis, the βA arms of three C subunits interact to form the β-annulus as in the native capsids In CP-P53A, the substitution of the conserved proline with alanine at position 53 does not seem to affect either the bending or the conformation of the β-annulus In the native as well as the recombinant T=3 capsids, there are fewer hydrogen bonds at the quasi 2-fold interface when compared to the icosahedral 2-fold interface mainly owing to the fact that the βA arms are ordered in the C subunits and hence are available for hydrogen bonding The structures of both rCP and CP-P53A capsids indicate that despite the differences in the nucleic acid that is encapsidated and the manipulation of the crucial proline at the β-annulus, the core of the CP and the interactions at various interfaces remain largely unaffected Hence, these elements might play little role in the assembly or capsid organization in SeMV

Chapter IV deals with the steps undertaken from crystallization to structure solution for the mutant T=1 capsids of SeMV CP-NΔ36 and CP-NΔ65 crystallized in space group $P2_1$, CP-NΔ65-CΔ2 crystallized in spacegroup the C2, CP-NΔ65-D146N-D149N crystallized in the spacegroup P1 and CP-NΔ31 crystallized in the spacegroup $P2_12_12_1$ with full virus particles in the crystallographic asymmetric unit X-ray data were collected for CP-NΔ31, CP-NΔ36, CP-NΔ65, CP-NΔ65-D146N-D149N to resolutions of 2 7 Å, 3 3 Å, 3 0 Å and 3 4 Å, respectively After collecting the X-ray diffraction data on the T=1 mutant capsids, rotation functions were calculated to establish the particle symmetry Self-rotation functions for κ values of 72°, 120° and 180° were calculated using the GLRF program These functions unambiguously established the particle symmetry and orientation In CP-NΔ65-D146N-D149N, the spacegroup was P1 and the origin was taken as the particle centre In the case of CP-NΔ36 and CP-NΔ65, a native Patterson map was calculated and the particle position suggested by a large peak observed in the Patterson function is x= 0 2406, z= 0 2542 while y=0 5 A densely packed T=1 capsid was constructed for using in molecular replacement trials by removing the hexamers followed by rotation and translation of the 12 pentamers The orientation and positions of the

model in the unit cell were optimized by varying their values by small increments and monitoring their effect on the crystallographic R_{free} factors in 20 cycles of rigid body refinement with CNS The parameters that resulted in the lowest R-value were then used for generating the 60 non-crystallographic symmetry matrices which were then used for refinement After getting the best possible orientation and position of the particle, the models were subjected to further refinement using "mlf" (maximum likelihood in amplitudes) option in CNS version 1 1 Complete data up to the highest resolution shell were used after setting aside 5% of the data for cross-validation for CP-NΔ31, CP-NΔ36 and CP-NΔ65 and 10% of the data for CP-NΔ65-D146N-D149N and CP-NΔ65-CΔ2 Sigma weighted $2F_o$-F_c and F_o-F_c maps were calculated and the model was built into them using the interactive model building program O The phases were improved using NCS averaging and solvent flattening carried out using locally developed programs for CP-NΔ36 and CP-NΔ65 The final model was built on to the NCS averaged map in these two cases

Chapter V deals with the structure analysis of mutant T=1 capsids of SeMV and their comparison with the native structure The structures of the deletion mutants retain accurately several key intersubunit interactions of the native virus The final model consists of 60 copies of CP Calcium ions are present in the inter-subunit interfaces in CP-NΔ31, CP-NΔ36 and CP-NΔ65 Calcium is, however, absent in CP-NΔ65-D146N-D149N and CP-NΔ65-CΔ2 The icosahedral pentameric subunits are exceedingly similar in the native and mutant structures Similarly, the dimeric structure of the T=1 capsids closely resembles the quasi dimers of the native structure The smaller curvature of the native particles results essentially from the occurrence of additional icosahedral dimers in the T=3 native structure T=1 structures of CP-NΔ31 and CP-NΔ36 show striking similarity to that of CP-NΔ65 Even in the former two, the density is prominent from only residue 73 The disorder of residues up to 72 in CP-NΔ31 and CP-NΔ36 may be due to the lack of the N-terminal residues that are required to induce order in the β-annulus segment In contrast to the close similarity in the structures of CP-NΔ31, CP-NΔ36 and CP-NΔ65, the structure of CP-NΔ65-D146N-D149N show larger changes Absence of

carboxylates which ligate calcium in CP-NΔ65-D146N-D149N leads to a structure which is expanded by as much as 2Å in comparison to the T=1 capsids of CP-NΔ31, CP-NΔ36 and CP-NΔ65 Another remarkable feature is lack of order at the C-terminus The carbonyl oxygen of N268 and the side chain OD of N267 are involved in calcium coordination in the native as well as in the T=1 capsids of CP-NΔ31, CP-NΔ36 and CP-NΔ65 Unlike the C-terminus, the region around residues 146 and 149, which corresponds to the sites of mutation, is not as disordered, presumably because of a compensatory interaction between OD1 of N146 and ND2 of N149 However, absence of calcium leads to large shifts in the CBS and the segments in its vicinity, which affects the subunit packing at the 3-fold axis leading to a large increase in contact distances Contacts across the 5-fold axis are also reduced, although to a lesser extent compared to 3-fold contacts

Chapter VI deals with the analysis of structural waters in CP-NΔ31 capsids The analysis is mainly concerned with probing whether the water molecules occur as integral parts of the viral CP and to address the contribution of water molecules to the intersubunit contacts at various interfaces Such water molecules are likely to contribute significantly to the stability of CP and packing of the subunits at different interfaces There are several completely buried water molecules in CP-NΔ31 suggesting that they constitute an integral part of the CP subunit They also contribute to the stability of loops and secondary structural elements within a subunit Many water molecules appear to be integral components of the intersubunit interfaces and contribute significantly to the inter-subunit contacts A comparative analysis of totally buried water molecules and water structures associated with loops in CP-NΔ31 and CP-NΔ65 has also been made The results suggest the importance of water molecules in subunit structure and capsid architecture

A part of the results presented in the thesis has been reported in the following publications

Sangita, V , Parthasarathy, S , Toma, S , Lokesh, G L , Gowri, T D , Satheshkumar, P S , Savithri, H S , and Murthy, M R (2002) Determination of the structure of the recombinant T=1 capsid of Sesbania mosaic virus *Curr Sci* **82(9),** 1123-1131